

AD_____

Award Number: W81XWH-07-1-0619

TITLE: Genetic and Environmental Pathways in Type 1 Diabetes
Complications

PRINCIPAL INVESTIGATOR: Massimo Trucco, M.D.

CONTRACTING ORGANIZATION: Children's Hospital of Pittsburgh
Pittsburgh, PA 15224

REPORT DATE: September 2009

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT:

X Approved for public release; distribution unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
<small>Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</small>					
1. REPORT DATE (DD-MM-YYYY) 01-09-2009		2. REPORT TYPE Annual		3. DATES COVERED (From - To) 27 August 2008 - 26 August 2009	
4. TITLE AND SUBTITLE Genetic and Environmental Pathways in Type 1 Diabetes Complications				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER W81XWH-07-1-0619	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Massimo Trucco, M.D.				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Children's Hospital of Pittsburgh Pittsburgh, PA 15224				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, MD 21702-5012				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution unlimited.					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT Genetic factors contribute to risk for developing nephropathy in patients with Type 1 Diabetes (T1D). Cigarette smoking is deleterious to kidney function and is a risk factor for Diabetic-Nephropathy (DN) as well as End Stage Renal Disease (ESRD) in patients with T1D. The proposed study investigates how environmental exposure(s) (e.g., smoking) and genetic variants interact to amplify risk for T1DN and substantially increase incidence of ESRD. The specific aims are: 1) Identify genetic variants conferring risk to T1DN by performing a staged follow-up of our initial Genome-Wide Association Scan (GWAS) results; 2) Ensure that SNPs identified by Aim 1 affect risk of T1DN, as opposed to risk for T1D; 3) Identify genetic variants that interact with smoking status in conferring risk for T1DN; 4) Confirm results obtained during Aims 1-3 using an independent cohort of case and control participants. The relevance of the study to public health is that 16 million people in the US have diabetes with 800,000 new cases diagnosed each year. Diabetic complications threatening vision, kidney, and nerve function affect most diabetic patients. Improved prediction of risk for developing diabetes and diabetic complications among active duty members of the military, their families and retired military personnel will potentially allow focused preventative treatment of at-risk individuals, providing significant healthcare savings and improved patient well being.					
15. SUBJECT TERMS End Stage Renal Disease; Genetic Association; Genome Scanning; Nephropathy; Type 1 Diabetes					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	18. NUMBER OF PAGES 23	19a. NAME OF RESPONSIBLE PERSON USAMRMC
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U			19b. TELEPHONE NUMBER (include area code)

**Children’s Hospital of Pittsburgh
W81XWH-07-1-0619
Annual Report (08/27/2008 – 08/26/2009)
Table of Contents**

Introduction	4
Body.....	
First Quarter.....	4
Second Quarter.....	9
Third Quarter.....	13
Fourth Quarter	20
Key Research Accomplishments.....	22
Conclusions.....	23
 Appendix	 N/A

INTRODUCTION:

Type 1 Diabetes (T1D) is associated with increased risk of T1D-Nephropathy (T1DN) and is usually accompanied by other diabetic-related complications such as retinopathy, neuropathy, blood pressure elevation, and high risk of cardiovascular morbidity and mortality. Sixteen million people in the US have diabetes with 800,000 new cases diagnosed each year. Diabetic complications affect most diabetic patients. Diabetes occurs in men, women, children and the elderly. African, Hispanic, Native and Asian Americans are particularly susceptible to its most severe complications. An estimated 20% to 40% of T1D patients will develop diabetic nephropathy, clinically first evidenced by microalbuminuria, during their lifetime. If untreated nearly all T1D patients experiencing microalbuminuria will progress to overt nephropathy, evidenced by macroalbuminuria, and culminating in T1D-End Stage Renal Disease (T1D-ESRD). Improved prediction of risk for developing diabetes and diabetic complications among active duty members of the military, their families and retired military personnel will potentially allow focused preventative treatment of at-risk individuals, providing significant healthcare savings and improved patient well being.

BODY:

Our first quarterly scientific progress report for the second year of our project (08/27/08 – 11/30/08) detailed the following steps forward in reaching the aims of our study.

The goals for the recently concluded research period were to: 1) expand the scope of the study to include the genetics of Type 1 Diabetic Nephropathy (T1DN); 2) genotype candidate SNPs for association with T1DN; and 3) analyze the results of SNP genotyping to determine which, if any, of the candidate SNPs resulted in a positive association signal. Our progress in the investigation of genetic signals correlating with T1DN is addressed below.

Goal 1. Expand genetic analysis to include the T1DN phenotype. **Milestone 1A.** Analyze available anthropometric data linked with the Type 1 Diabetes (T1D) cohort for presence of diabetes complications. **Milestone 1B.** Determine the size of the available cohort for T1DN and T1D with healthy kidney function.

Completion of Milestone 1A: Analysis of the available anthropometric data linked with the T1D cohort for presence of diabetes complications has been completed. Cohorts with T1D-ESRD or T1DN (treated as cases), or T1D (without DN, treated as controls) were used as a resource to evaluate the genetics of T1DN. For Stage 1 of the study 197 participants with T1D-ESRD and reporting never having smoked cigarettes were chosen as cases (Table 1). Control samples (N=197) were chosen from among the T1D cohort that reported never having smoked cigarettes and having at least 15 years duration of T1D and no evidence of DN.

The Stage 2 cohort was also chosen from among the overall study population. It includes N=634 cases comprised of a mixture of T1D-ESRD and T1DN participants. The control population (N=711) included T1D participants that mirror the characteristics observed in the Stage 1 cohort (Table 1). However, both cases and controls were allowed to include smokers as well as participants reporting never having smoked cigarettes. Cigarette smoking has been implicated as an influential factor leading to increased incidence of kidney disease. Inclusion of smokers in the Stage 2 cohort enabled analysis of gene and environment interaction for select SNPs by using smoking as a covariant during analysis.

Table 1. Combined study cohort clinical data.

	Stage 1 Singleton Cohort		Stage 2 Singleton Cohort	
	Case (N=197)	Control (N=197)	Case (N=634)	Control (N=711)
<i>Demographic Characteristics:</i>				
Female	105(53%)	105(53%)	334(53%)	276(39%)
Male	92(47%)	92(47%)	300(47%)	435(61%)
Smoking	0	0	411(65%)	293(41%)
Body Mass Index (kg/m ²)	26±6	26±4	26±5	26±4
Age at Enrollment (yrs)	45±6	42±8	42±7	37±9
<i>History of Diabetes:</i>				
Age at T1D Diagnosis (yrs)	12±7	12±8	12±7	13±7
Duration of T1D (yrs)	33±7	30±9	31±8	24±7
Pancreas Transplant	107(54%)	0	158(25%)	0
HbA1C (%) with Pancreas Transplant	5.8±1.7	---	5.7±1.3	---
HbA1C (%) without Pancreas Transplant	7.9±1.5	7.3±1.1	8.4±1.6	7.5±1.2

History of Nephropathy:

T1DN	0	0	287(45%)	0
T1D-ESRD	197(100%)	0	347(55%)	0
Time to Onset of T1D-ESRD (yrs)	26±8	----	25±7	----
Serum Creatinine (mg/dL) ¹	----	0.9±0.2	1.8±1.0	0.9±0.2
Serum Cystatin C (mg/L) ¹	----	0.8±0.1	1.8±1.0	0.8±0.1
eGFR (ml/min/1.73m ²) ¹	----	85±16	52±25	88±17

1. Serum creatinine, serum cystatin C, and eGFR are reported for the T1DN and T1D participants only.

Completion of Milestone 1B: Determination of the size of the available cohort for T1DN and T1D with healthy kidney function has been completed. The number of participants available for the T1DN case versus T1D control analysis is listed in Table 2. The participants differ in the fraction of men and women available for the study and in the fraction of participants who reported having smoked cigarettes. An important attribute of the study population is that it includes a substantial number of participants (i.e., 823 cases and 912 controls) that can be used to investigate genetic signals for their association with the T1DN phenotype.

Table 2. Summary of the Combined Cohorts. T1D-ESRD/T1D (Smokers/Non-Smokers) and T1D (Smokers/Non-Smokers)

	<u>Case</u>	<u>Control</u>
Total	823	912
Female	405(49%)	541(59%)
Male	427(52%)	371(41%)
T1D-ESRD	547(66%)	---
Smoking	411(50%)	298(33%)

Goal 2. Garner genetic data to test for association of select SNPs with T1DN. **Milestone 2A.** Use the T1DN and T1D cohort identified during Goal 1 to initiate genotyping of candidate SNPs for their association with T1DN.

Completion of Milestone 2A: Use of the T1DN and T1D cohort identified during Goal 1 to genotype candidate SNPs for their association with T1DN has been initiated. Figure 1 summarizes results obtained from the GWA scan obtained using the Stage 1 cohort. The figure indicates the association for individual SNPs with the T1DN-ESRD phenotype for the additive mode of inheritance. The strongest observed p-value occurred at rs2215930 on Chromosome 19 within the *ZNF71* locus and was 5.7×10^{-7} .

Other genomic regions exhibiting SNPs with relative small p-values (i.e., those among the smallest 15 p-values observed for the additive, dominant, or recessive modes of inheritance) are listed in Table 3. Of these strong signals 9 occurred within 7 genes (*THADA*, *FLNB*, *CTNND2*, *ZNF804B*, *SLC1A2*, and *ZNF71*) and 11 were intergenic. Nine of the 11 occurred within 14kb and 300kb of known loci, and 2 (rs921186 and rs11763265) occurred at greater than 300kb from the nearest gene.

Of concern for GWA studies is whether the association signals are due to poor genotyping calls. Evidence against this possibility comes from SNPs with correlated genotypes that due to linkage disequilibrium (LD) show similar patterns of association. Of the 18 genomic regions and SNPs showing strong association signals, 11 correlated with 3 or more SNPs occurring within 50kb of one another exhibiting p-values less than 0.001 (data not shown). The largest blocks of correlated SNPs occurred within the *SLC1A2* locus (rs11033073), 12 SNPs; on Chromosome 13q33.3 centromeric to the *IRS2* locus (rs9521445), 8 SNPs; and Chromosome 9q22.33 telomeric to the *SEC61B* locus (rs7040144), 7 SNPs. Comparison of GWA results with LD data obtained from HapMap, Build 36 reduced the number of SNPs in each group by informing the choice of tag-SNPs (Table 3). The tag-SNPs were selected using a threshold of $r^2 \leq 0.8$ in order to effectively reduce the number of SNPs needed for initial replication of the GWA results. Because all SNPs in Table 3 showed reasonable clustering of genotypes, and most have correlated SNPs showing similar patterns of association, the 21 SNPs in Table 3 were used to evaluate the results observed during the Stage 1 GWA scan.

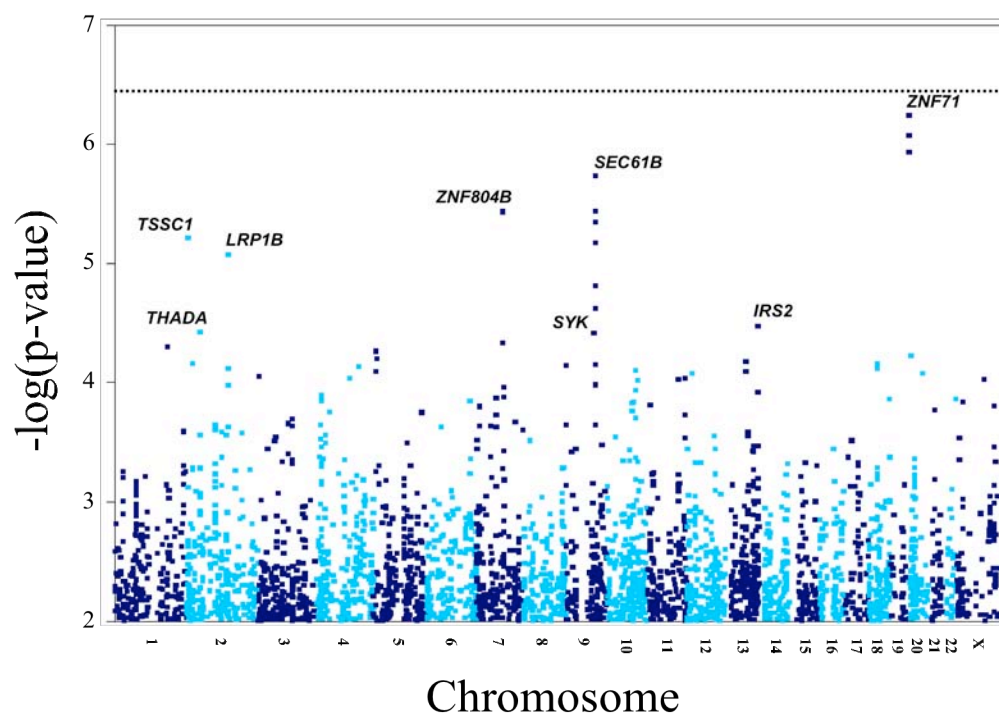


Figure 1. Genome location versus $-\log(p\text{-value})$ for the T1D-ESRD (case) and T1D (control) genome-wide scan. Genotyping data from 322,347 SNPs were used to determine the p-values associated with the additive mode of inheritance. The Bonferroni corrected p-value was estimated at 3.6×10^{-7} for genome-wide significance at $p < 0.5$ and is indicated by a horizontal dotted line.

Table 3. Summary of strong association signals obtained from the Stage 1 cohort.

dbSNP ID	Chr	Location	Locus	Genotype Based Typing (p-value)*		
				Additive	Dominant	Recessive
rs1470614	2	2,893,299	<i>TSSC1</i>	6.2×10^{-6}	6.4×10^{-6}	
rs11693666	2	43,415,983	<i>THADA</i>	3.8×10^{-5}	8.1×10^{-5}	
rs7586146	2	140,489,344	<i>LRP1B</i>	8.5×10^{-6}	1.5×10^{-5}	
rs1718460	3	57,981,453	<i>FLNB</i>		7.5×10^{-5}	
rs842523	3	153,242,605	<i>SUCNR1</i>			2.6×10^{-5}
rs921186	5	8,521,123	Gene Poor Region	5.5×10^{-5}		1.1×10^{-5}
rs26456	5	11,316,195	<i>CTNND2</i>	6.4×10^{-5}		4.3×10^{-5}
rs4867592	5	169,580,409	<i>LCP2</i>		1.6×10^{-5}	
rs11763265	7	66,625,751	Gene Poor Region	1.4×10^{-4}		4.6×10^{-5}
rs9640583	7	88,647,585	<i>ZNF804B</i>	3.7×10^{-6}		2.2×10^{-5}
rs2214339	7	88,651,225	<i>ZNF804B</i>	4.7×10^{-5}	1.8×10^{-5}	
rs2184540	9	92,841,029	<i>SYK</i>	3.8×10^{-5}	3.2×10^{-4}	
rs7040144	9	101,216,224	<i>SEC61B</i>	1.9×10^{-6}	1.1×10^{-4}	3.7×10^{-5}
rs4743350	9	101,217,469	<i>SEC61B</i>	3.7×10^{-6}		2.9×10^{-5}
rs11033073	11	35,310,520	<i>SLC1A2</i>	9.4×10^{-4}	2.3×10^{-5}	
rs12365943	11	108,608,422	<i>C11orf87</i>	9.5×10^{-5}	1.9×10^{-6}	
rs11222530	11	130,599,312	<i>HNT</i>	9.2×10^{-5}	4.1×10^{-5}	
rs9521445	13	109,083,535	<i>IRS2</i>	3.3×10^{-5}	3.0×10^{-4}	
rs11656096	17	19,563,235	<i>SLC47A2</i>	3.1×10^{-4}	2.1×10^{-5}	
rs11673097	19	61,811,246	<i>ZNF71</i>	8.5×10^{-7}		4.4×10^{-6}
rs2215930	19	61,814,858	<i>ZNF71</i>	5.7×10^{-7}	1.9×10^{-5}	1.6×10^{-4}

*Only p-values less than 0.001 are indicated. At least one p-value was among the top 15 signals observed for the indicated modes of inheritance.

†Distance between SNP and nearest locus if intergenic. If SNP is intragenic then its functional position is indicated.

Goal 3. Analyze the results of genotyping obtained using the T1DN cohort. **Milestone 3A.** Using T1DN family trios calculate the p-value for association of candidate SNPs with the T1DN phenotype. **Milestone 3B.** Using T1DN and T1D singletons calculate the p-value for association of select SNPs with the T1DN phenotype.

Completion of Milestone 3A: Using T1DN family trios calculate the p-value for association of candidate SNPs with the T1DN phenotype. Of the 21 SNPs in Table 4 all but two (rs1470614 and rs2184540) could be genotyped using the platforms we had available. Failure to genotype rs1470614 and rs2184540 were due to poor signal to noise observed during development of the TaqMan as well as iPLEX based SNP typing assays (data not shown). Thus 19 SNPs were genotyped in Stage 2. An initial portion of the replication phase genotyped N=136 affected family trios.

Transmission disequilibrium test (TDT) analysis of genetic data obtained from the combined T1DN and T1D-ESRD affected family trios indicated that none of the markers achieved an uncorrected p-value exceeding the p-value threshold of 0.05 (Table 4). This result was expected due to the relatively small genetic risk ratio (2 to 3-fold) associated with T1DN and T1D-ESRD and the limited power of this trio set. For this reason, TDT data we used to prioritize SNPs for subsequent replication stage analysis using the Stage 2 cohort of N=634 cases and N=711 controls (Table 1). This was accomplished by grouping potentially positive SNPs, identified during analysis of the Stage 1 GWA scan and screened by TDT analysis, into two classes based upon whether transmission of the more frequently inherited allele occurred within the third quartile for all transmissions ($\geq 57\%$). As illustrated in Table 4, there are six SNPs (i.e., rs7586146, rs26456, rs11763265, rs2214339, rs9521445, and rs11673097) that showed modest evidence for biased transmission ($\geq 57\%$).

Table 4. TDT analysis of T1DN and T1D-ESRD affected family trios (N=136).

dbSNP ID*	Chr	Location	Locus	Number of Transmissions			Chi Sq	p-value	Percent
				Allele 1	Allele 2	Total			Trans†
Class 1 (% transmission exceeded the 3rd quartile):									
rs7586146	2	140,489,344	LRP1B	22	16	38	0.9	0.33	58
rs26456	5	11,316,195	CTNND2	18	25	43	1.1	0.29	58
rs11763265	7	66,625,751		33	48	81	2.8	0.1	59
rs2214339	7	88,651,225	ZNF804B	30	40	70	1.4	0.23	57
rs9521445	13	109,083,535	IRS2	62	47	109	2.1	0.15	57
rs11673097	19	61,811,246	ZNF71	46	62	108	2.4	0.12	57
Class 2 (% transmission did not exceed the 3rd quartile):									
rs1470614	2	2,893,299	TSSC1	----	----	----	----	----	----
rs11693666	2	43,415,983	THADA	42	46	88	0.2	0.67	52
rs1718460	3	57,981,453	FLNB	51	60	111	0.7	0.39	54
rs842523	3	153,242,605	SUCNR1	59	61	120	0	0.86	51
rs921186	5	8,521,123		38	45	83	0.6	0.44	54
rs4867592	5	169,580,409	LCP2	47	54	101	0.5	0.49	53
rs9640583	7	88,647,585	ZNF804B	31	36	67	0.4	0.54	54
rs2184540	9	92,841,029	SYK	----	----	----	----	----	----
rs7040144	9	101,216,224	SEC61B	47	52	99	0.3	0.62	53
rs4743350	9	101,217,469	SEC61B	45	44	89	0	0.92	51
rs11033073	11	35,310,520	SLC1A2	61	50	111	1.1	0.3	55
rs12365943	11	108,608,422	C11orf87	53	52	105	0	0.92	50
rs11222530	11	130,599,312	HNT	36	34	70	0.1	0.81	51
rs11656096	17	19,563,235	SLC47A2	57	53	110	0.1	0.7	52
rs2215930	19	61,814,858	ZNF71	57	65	122	0.5	0.47	53

*Data was not reported for 2 SNPs (i.e., rs1470614 and rs2184540) due to poor genotyping quality.

†Percent transmission of the more highly inherited allele was used to determine the 3rd quartile for transmission, i.e., 57%.

Completion of Milestone 3B: Using T1DN and T1D singletons calculate the p-value for association of select SNPs with the T1DN phenotype. The six SNPs identified in Table 4 as occurring within the third quartile for transmission were examined using the combined Stage 1 and 2 cohorts of cases and controls. The results indicated that SNP rs9521445, occurring at 13q33.3, exhibited genotyping data consistent with association for the T1DN phenotype ($p=1.8 \times 10^{-3}$) while p-values exhibited by the other markers failed to fall below the threshold of $p < 0.05$ (Table 5).

LD data available for SNPs identified during the GWA scan were used to determine a set of tag-SNPs that would enable fine mapping of the 13q33.3 region surrounding rs9521445. SNPs were chosen among those occurring between rs12864280 and rs9521471, representing a roughly 100kb region on Chromosome 13, in order to capture the peak signal identified by the Stage 1 GWA results. The results from fine mapping showed two SNPs chosen as tags, namely rs17412858 and rs7989418, and another SNP, rs2150479 (modestly in LD with rs17412858), resulted in p-values for association that fall below 0.006. The strongest signal was associated with rs7989418 and was 2.6×10^{-4} with an odds ratio of 1.7 (95%CI, 1.3-2.3). The SNP rs7989418, a tag-SNP, tags a cluster that also contains rs9521445, the SNP initially identified during the GWA scan. These two SNPs are substantially correlated ($r^2=0.70$).

Table 5. Summary of Cohort 1 genotyping results for fine-mapping on Chromosome 13q33.3.

dbSNP ID*	Location	Genotype Based Typing (p-value)			OR(95%CI)
		Additive	Dominant	Recessive	
rs7330849	109,034,647	2.5×10^{-2}	5.1×10^{-2}	7.4×10^{-2}	1.4(1.1-2.0)
rs4462453	109,049,329	9.6×10^{-1}	9.7×10^{-1}	9.6×10^{-1}	1.0(0.7-1.4)
rs17412858	109,050,609	1.0×10^{-2}	6.0×10^{-3}	2.2×10^{-1}	1.5(1.1-2.1)
rs331628	109,068,721	8.2×10^{-1}	1	8.2×10^{-1}	1.1(0.4-2.8)
rs7989418	109,081,443	2.6×10^{-4}	1.8×10^{-3}	5.1×10^{-3}	1.7(1.3-2.3)
rs9521446	109,086,420	3.2×10^{-2}	2.4×10^{-2}	2.8×10^{-1}	1.4(1.0-1.8)
rs1330547	109,100,842	2.1×10^{-1}	2.7×10^{-1}	3.9×10^{-1}	1.2(0.9-1.8)
rs9515094	109,107,932	5.6×10^{-1}	6.7×10^{-1}	6.1×10^{-1}	1.1(0.8-1.5)
rs1411554	109,108,321	1.8×10^{-2}	1	1.8×10^{-2}	1.9(1.1-3.4)
rs149577	109,116,649	6.7×10^{-1}	1.6×10^{-1}	4.5×10^{-1}	1.1(0.6-2.0)
rs9521467	109,125,704	6.6×10^{-1}	5.1×10^{-1}	8.1×10^{-1}	1.1(0.8-1.6)
rs9521468	109,127,880	3.9×10^{-1}	5.4×10^{-1}	4.3×10^{-1}	1.1(0.8-1.5)
rs9521469	109,128,522	1.9×10^{-2}	1.5×10^{-1}	2.3×10^{-2}	1.7(1.1-2.5)
rs996969	109,040,933	6.8×10^{-2}	2.4×10^{-2}	1.8×10^{-1}	1.7(1.0-2.9)
rs2391776	109,041,426	2.8×10^{-1}	7.9×10^{-1}	1.7×10^{-1}	1.2(0.8-1.8)
rs2150479	109,041,870	1.7×10^{-2}	3.6×10^{-3}	6.9×10^{-1}	1.5(1.1-2.1)
rs11069790	109,042,402	5.3×10^{-2}	9.8×10^{-2}	1.0×10^{-1}	1.6(1.0-2.5)
rs9521439	109,058,768	6.0×10^{-2}	1.1×10^{-1}	4.9×10^{-2}	1.8(1.1-2.8)
rs17504752	109,088,096	7.7×10^{-2}	5.1×10^{-2}	1	1.4(0.9-2.0)
rs184547	109,090,545	1.0×10^{-1}	1	1.0×10^{-1}	2.5(0.8-8.2)
rs9515093	109,107,664	3.2×10^{-1}	7.0×10^{-1}	2.2×10^{-1}	1.3(0.8-1.9)
rs12875288	109,127,498	6.7×10^{-1}	6.0×10^{-1}	1	1.1(0.8-1.5)
rs7337197	109,128,614	2.7×10^{-1}	1.6×10^{-1}	4.4×10^{-1}	1.3(0.8-1.9)

*Bold font indicates SNPs with $p < 0.006$ for at least one mode of inheritance (e.g., additive, dominant, or recessive).

Statement of Plans for the upcoming research period

Goal 1. Evaluate genotyping results from fine mapping of the Chromosome 13q33.3 region for covariance with intrinsic characteristics of the study cohort. **Milestone 1A.** Determine whether the signal associated with T1DN varies with gender or age at onset of T1D.

Goal 2. Evaluate genotyping results from fine mapping of the Chromosome 13q33.3 region for covariance with exposure to environmental variables. **Milestone 2A.** Determine whether the signal associated with T1DN varies with smoking history as well as clinical measurements of renal status, duration of T1D prior to onset of ESRD, or body mass.

Goal 3. Organize the study results for inclusion into a publication describing the interaction between and environment and susceptibility for development T1DN. **Milestone 3A.** Prepare a manuscript describing the overall study aims and results. **Milestone 3B.** Submit the manuscript for publication in a peer-reviewed journal.

In our second quarterly scientific progress report (12/01/08 – 02/28/09), we presented the following data:

There were 3 research goals to have been addressed during the last research quarter: 1) evaluate genotyping results from fine mapping of the Chromosome 13q33.3 region for covariance with intrinsic characteristics of the study cohort; 2) evaluate genotyping results from fine mapping of the Chromosome 13q33.3 region for covariance with exposure to environmental variables; and 3) organize the study results for inclusion into a publication describing the interaction between and environment and susceptibility for development T1DN. The goals and milestones have been completed and the experimental data is summarized below.

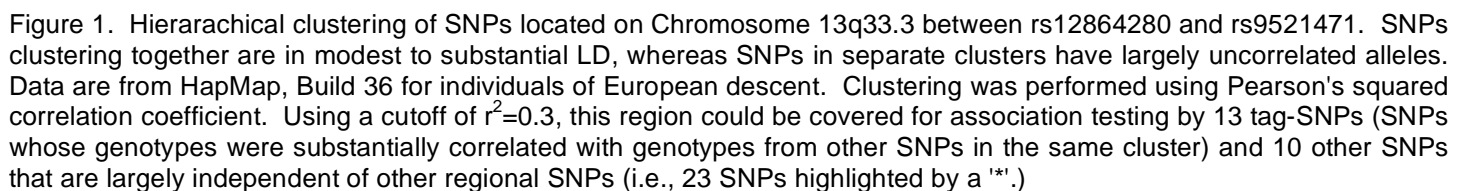
Goal 1. Evaluate genotyping results from fine mapping of the Chromosome 13q33.3 region for covariance with intrinsic characteristics of the study cohort. **Milestone 1A.** Determine whether the signal associated with T1DN varies with gender or age at onset of T1D.

Completion of Milestone 1A: From the results presented in the preceding Quarterly Report we identified 3 SNPs (rs17412858, rs7989418, and rs2150479) located on Chromosome 13q33.3 with compelling evidence for association with T1DN. Using the information collected on these SNPs as a foundation for our study we broadened coverage by selecting 3 other SNPs in modest linkage disequilibrium (LD) with the original SNPs: specifically we included rs1547241, rs1041466, and rs2391777 (Figure 1). The data summarized in Table 1 for the combined cohort indicated that the strongest signals were associated with rs1041466 and rs2391777 all falling in one cluster (Figure 1); with rs17412858 and rs7989418 falling in another cluster; as well as rs2150479, which is not materially correlated with the other SNPs. Analyses including gender indicate that association between the Chromosome 13q33.3 region and the phenotype could be different between men and women, the sub-sample containing only women showed the stronger association with p-values $<10^{-5}$ occurring for rs17412858, rs7989418, and rs2150479 (Table 1, middle set of data).

Table 1. Genotyping results for fine mapping Chromosome 13q33.3 using the combined cohort.

tag-SNP	Genotype Based Typing (p-value)			OR(95%CI)
	Additive	Dominant	Recessive	
Combined Cohort 1 and 2:				
rs17412858	1.8x10 ⁻⁵	1.7x10 ⁻³	3.7x10 ⁻⁵	1.4(1.2-1.6)
rs1547241	6.9x10 ⁻³	2.3x10 ⁻²	2.9x10 ⁻²	1.2(1.1-1.4)
rs1041466	2.9x10 ⁻⁶	3.5x10 ⁻⁴	3.4x10 ⁻⁵	1.4(1.2-1.6)
rs2391777	1.9x10 ⁻⁵	2.0x10 ⁻²	1.0x10 ⁻⁵	1.4(1.2-1.6)
rs7989418	1.4x10 ⁻⁵	5.8x10 ⁻³	1.6x10 ⁻⁵	1.4(1.2-1.6)
rs2150479	2.6x10 ⁻⁶	7.9x10 ⁻⁵	1.2x10 ⁻⁴	1.5(1.3-1.7)
Combined Cohort 1 and 2 Females:				
rs17412858	2.1x10 ⁻⁶	8.3x10 ⁻⁴	1.9x10 ⁻⁶	1.5(1.3-1.8)
rs1547241	2.9x10 ⁻⁴	3.7x10 ⁻³	1.8x10 ⁻³	1.4(1.2-1.7)
rs1041466	6.0x10 ⁻⁴	8.3x10 ⁻⁴	1.8x10 ⁻²	1.3(1.1-1.6)
rs2391777	6.7x10 ⁻⁵	2.5x10 ⁻²	5.0x10 ⁻⁵	1.4(1.2-1.7)
rs7989418	1.0x10 ⁻⁶	3.5x10 ⁻³	4.4x10 ⁻⁷	1.5(1.3-1.8)
rs2150479	4.5x10 ⁻⁷	4.8x10 ⁻⁵	1.0x10 ⁻⁵	1.6(1.3-1.9)
Combined Cohort 1 and 2 Female Smokers:				
rs17412858	3.0x10 ⁻⁵	7.8x10 ⁻³	3.8x10 ⁻⁶	1.7(1.3-2.1)
rs1547241	2.5x10 ⁻³	3.8x10 ⁻²	2.0x10 ⁻³	1.4(1.1-1.8)

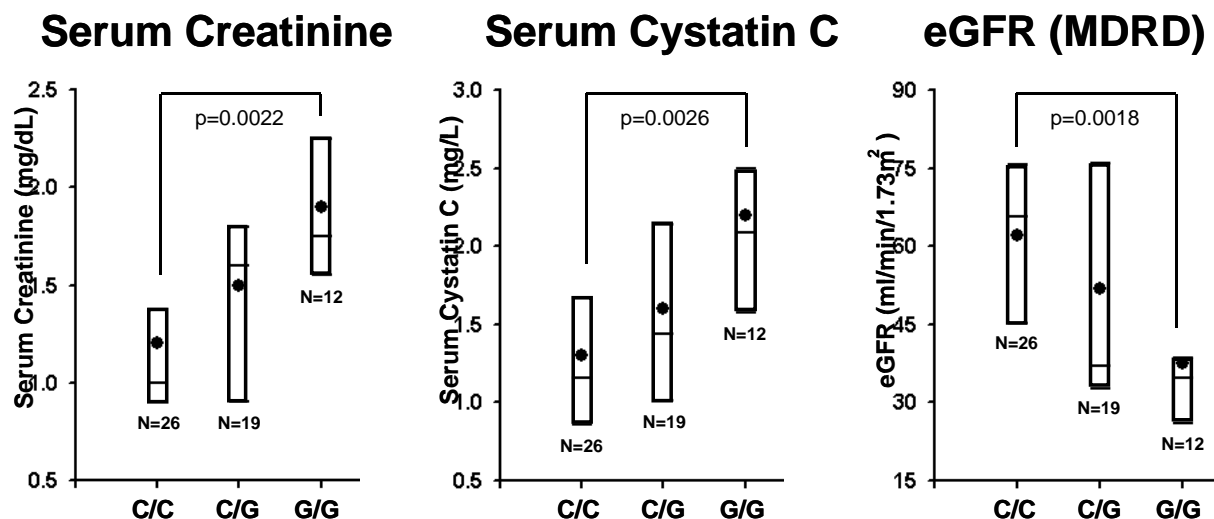
Hierarchical Clustering Dendrogram: rs12864280–rs9521471



Completion of Milestone 2A: Experimental results obtained from genotyping SNPs located in the Chromosome 13q33.3 region were examined to test the hypothesis that clinical measures of severity of disease would co-vary. Evaluation of environment and gene interactions (Table 1, lower set of data) indicated that exposure to cigarette smoking in women cases (N=179) and controls (N=908) was associated with the rs2150479-G allele exhibiting a p-value exceeding 2.1×10^{-7} for a recessive mode of inheritance with an odds ratio of 1.8 (95% CI 1.4-2.3).

Clinical data from the combined cohort included measurements of serum creatinine and serum cystatin C that were obtained near the time of enrollment into the study. Anthropometric data included the age of the participant at enrollment and allowed an estimate of the severity of diabetic nephropathy by enabling the calculation of the glomerular filtration rate (eGFR). As illustrated in Figure 2, among women with T1DN who have reported smoking cigarettes (N=57) those with rs2150479 G/G genotype exhibited increased serum creatinine ($p=0.0022$) and serum cystatin C ($p=0.0026$), as well as decreased eGFR ($p=0.0018$) consisting of an average eGFR equal to 38 ml/min/1.73m² and consistent with severe clinical manifestation of T1DN. Likewise, in the T1D-ESRD cohort of women smokers (N=116) examination of the duration of T1D prior to occurrence of ESRD indicated that the rs2150479 G/G genotype associated with earlier occurrence of ESRD, as determined by treatment with dialysis or receipt of kidney transplantation ($p=0.015$), occurring at 25.5 years for G/G genotype versus 32 years duration for T1D for 80% incidence in the combined cohort (Figure 3).

Figure 2. The relationship between genotype for rs2150479 and serum creatinine, serum cystatin C, and eGFR



measured in T1DN females who reported having smoked cigarettes (N=57). Genotypes are shown along the x-axis. Boxes indicate the first and third quartile, medians are marked by the horizontal line, and circles denote mean values. The eGFR was calculated from serum creatinine levels measured near the time of enrollment into the study and used the relationship established by the Modification of Diet in Renal Disease (MDRD) study that $eGFR = 186 \times \text{serum creatinine (mg/dL)}^{-1.154} \times \text{age}^{-0.203} \times 0.742$ (27). The p-value calculated using the Mann-Whitney statistic is indicated for a comparison of the rs2150479 C/C (N=26) and G/G (N=12) genotypes, $p=0.0022$ (serum creatinine), $p=0.0026$ (serum cystatin C), and $p=0.0018$ (eGFR).

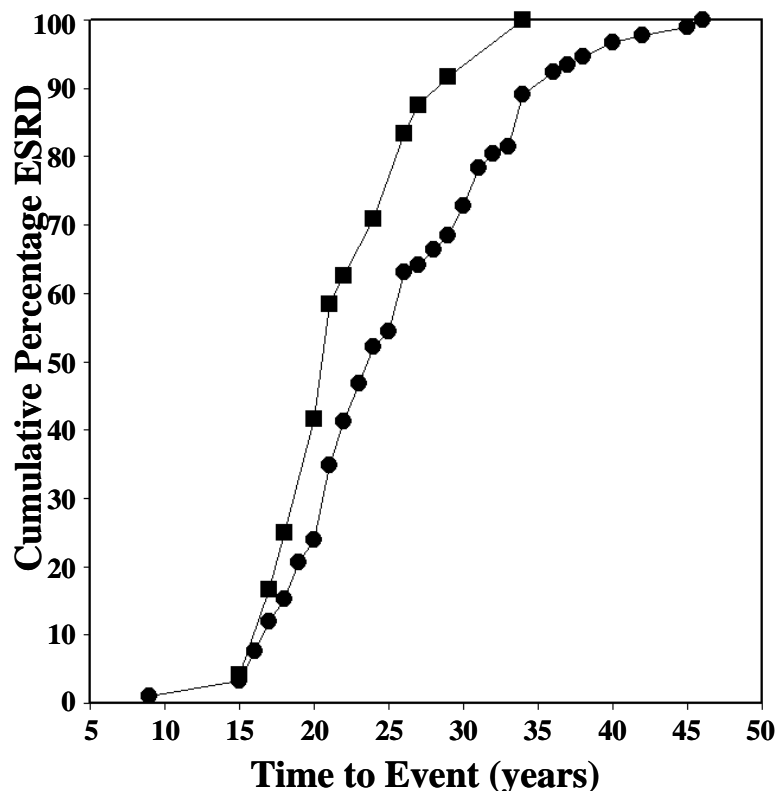


Figure 3. Time of duration from onset of T1D to occurrence of ESRD measured in T1D-ESRD females who reported having smoked cigarettes (N=116). SNP rs2150479 genotypes C/C plus C/G (N=92), circles, and G/G (N=24), squares are indicated. Onset of ESRD was defined as date of treatment by dialysis or receipt of a kidney transplant whichever occurred first. The p-value was calculated using the log rank test and indicated a p-value=0.015.

Goal 3. Organize the study results for inclusion into a publication describing the interaction between and environment and susceptibility for development T1DN. **Milestone 3A.** Prepare a manuscript describing the overall study aims and results. **Milestone 3B.** Submit the manuscript for publication in a peer-reviewed journal.

Completion of Milestone 3A: A manuscript summarizing the results of the study has been prepared. The study examined results from a GWAS in order to identify SNPs associated with T1DN. Performed in two stages the study evaluated genetic variants using DNA obtained from cases and controls. The results identified a region on Chromosome 13q33.3 distal to the *IRS2* locus. SNPs in this genomic region showed a p-value of 2.6×10^{-6} for the SNP rs2150479 when the complete cohort of N=831 cases and N=908 controls were compared. Evaluation of rs2150479 using gender and smoking as co-variant indicate a p-value of 2.1×10^{-7} when female smoker T1DN cases are examined. Moreover, comparison of T1D duration with incidence of ESRD indicated that participants with rs2150479 G/G genotype experienced ESRD after shorter duration times. The principal conclusions are that a region on Chromosome 13q33.3 is associated with T1DN and that women who have smoked and carry rs2150479 genotype G/G are more likely to suffer reduced kidney function and increased risk for ESRD.

Completion of Milestone 3B: The manuscript has been submitted to the journal DIABETES for publication. A copy of the manuscript is included as an Appendix to this quarterly report.

12. Statement of Plans for the upcoming research period

Goal 1. Extend genetic evaluation of T1DN by incorporating the results from GWAS published in the scientific literature. **Milestone 1A.** Incorporate GWAS data gathered by GAIN. **Milestone 1B.** Incorporate GWAS of data garnered by DCCT/EDIC.

Goal 2. Using the results from the combined analysis of Children's Hospital of Pittsburgh and publicly available data choose genomic regions for fine mapping. **Milestone 2A.** Select regions for fine mapping. **Milestone 2B.** Select tag-SNPs that will be genotyped. **Milestone 2C.** Design experimental assays for laboratory analysis.

Goal 3. Initiate fine mapping of T1DN susceptibility regions. **Milestone 3A.** Genotype selected SNPs in T1DN cases and T1D controls. **Milestone 3B.** Analyze results to determine association between variant(s) and T1DN phenotype.

In our third quarterly scientific progress report (03/01/09 – 05/31/09) we then reported the following findings:

At an early time point during the recently completed research quarter we became aware of 2 opportunities to expand the productivity of the project. Namely, these were firstly to publish data obtained from the research and secondly to develop a collaboration with a proteomics center. An outcome of this decision was that the goals as they were defined at the end of the last quarter's research report have been postponed. These goals are now the emphasis of the upcoming research period and are listed at the end of this report under the section heading Statement of Plans for the Upcoming Research Period.

The 2 research goals for the recently completed research period were: 1) write and submit a manuscript to a scientific peer-reviewed journal describing our work developing methods for analyzing genetic signals; and 2) expand the focus of our research to include proteomics research designed to quantify proteins in circulation that can be tied first to onset of T1D and then to risk of developing diabetes complications. The goals and milestones stated in the opening paragraph have been completed and the experimental data is summarized below.

Goal 1. Write and submit a manuscript to a scientific peer-review journal describing our work developing methods for analyzing genetic signals.

See Appendix A for a preprint of the manuscript. The paper was submitted to the journal Genetic Epidemiology on June 8, 2009.

Goal 2. Expand the focus of our research to include proteomics research designed to quantify proteins in circulation that can be tied first to onset of T1D and then to risk of developing diabetes complications.

BIOMARKER DISCOVERY FOR IMPROVED PREDICTION OF TYPE 1 DIABETES

Goal 2. SUMMARY

Type 1 Diabetes (T1D) is one of the most common chronic diseases of childhood. The project goal is to discover biomarkers present in serum that can be used to predict risk of developing T1D. The discovery of biomarkers that are highly sensitive and specific are critical for identifying at risk populations and will aid in the creation of new therapeutic treatments.

Goal 2. BACKGROUND

In patients with T1D there is T-cell mediated autoimmune destruction of pancreatic beta cells that results in loss of insulin production (Trucco, 2005; Pasquali et al., 2006). Onset of diabetes is commonly believed to occur when greater than 90% of the beta cell mass has been depleted (Barker and Eisenbarth, 2004). A large fraction of risk for developing the disease is genetic (estimated at between 30% to 50%) and environmental influences such as exposure to infectious agents have been suggested to account for the remainder (Redondo et al., 2008; Filippi and von Herrath, 2008). Measurement of plasma insulin or C-peptide levels can be used to estimate functional capacity of beta cells (Steele et al., 2004). However, using these methods to quantify the functional beta cell mass is not possible as illustrated by studies showing that beta cells are capable of increased insulin production to meet metabolic demands (Butler et al., 2007). At present a substantial number of T1D patients are diagnosed upon sudden illness and frequently exhibit severe dehydration, ketoacidosis, and hyperglycemia (Rewers et al., 2002; Barker and Eisenbarth, 2004).

T1D primarily strikes children and young adults (less than 20 years of age) with greater than 30,000 new cases occurring annually in the US (Karvonen et al., 2000). The incidence of T1D in the U.S. is between 10 and 15

cases per 100,000 per year and the number of people in the US living with T1D is estimated between 1 to 1.4 million (Newacheck and Taylor, 1992; Karvonen et al., 2000). Screening relatives of T1D patients for autoantibodies is capable of providing reliable risk estimation of T1D; assessment of risk can be performed by measuring for the presence of autoreactive antibodies to glutamic acid decarboxylase (GAD65), insulinoma-associated antigen-2 (IA-2), insulin (IAA), or islet cell antigen (ICA) (Riley et al., 1990; Pietropaolo et al., 2005). Evidence supporting this approach as a reliable surrogate to identify at risk individuals was provided by the Diabetes Prevention Trial (DPT-1) in a cohort of first-degree relatives of T1D patients. The DPT-1 observed that individuals with 2 or more autoantibodies exhibited an overall 68% 5 years risk while those with 3 autoantibodies approached a near absolute risk for developing T1D (Verge et al., 1996). However, the population of T1D patients with affected first-degree relatives accounts for less than 15% of new T1D cases (Pihoker et al., 2005). In the US population roughly 5% are anticipated to exhibit anti-islet antibodies, a number nearly 50-fold greater than the estimated number of individuals living with T1D. Screening for genetic risk alleles (e.g., HLA genotype) in first-degree relatives is possible, however, as was observed for autoreactive antibodies HLA genotype is a poor predictor of T1D risk in the general population. High-risk HLA DRB1-DQA1-DQB1 haplotypes 0301-0501-0201 and 0401-0301-0302 occur in roughly 17.3% of the European-American population (Klitz et al., 2003). As was observed for risk assessment based upon the presence of autoreactive antibodies the occurrence of high-risk HLA haplotypes greatly exceeds the prevalence of T1D in the population.

Diseases such as T1D in which symptoms occur late in disease progression, and typically exhibit slow progress of disease, provide an opportunity for diagnosis before clinical symptoms occur. Among the reasons to develop new diagnostic tools for prediction of T1D risk are the need to enable early diagnosis of the disease and to assess efficacy of therapeutic interventions. Early detection of T1D risk may allow stratification of cohorts prior to assessment of treatments that improve understanding of the underlying pathogenesis. Accurate diagnosis can aid in selection of patients likely to respond to new therapies as well as provide a means to stratify participants in clinical trials.

Diagnosis of individuals at risk for T1D can improve our insight into changes associated with disease onset and assessment of therapies to reverse diabetes. For example, in clinical trials performed to date autoantibodies have not correlated with intervention outcomes affecting disease progression (Palmer et al., 2004). Development of methodologies that enable stratification of patients by risk can benefit patient populations by enabling the rationale use of treatment options. Screening of children for T1D risk may be used when selecting patients for clinical trials in which therapies designed to prevent disease are being applied. Criteria for identifying candidate diagnostic biomarkers are: 1) that they are sensitive being able to identify at risk persons and 2) that they are specific being capable of distinguishing between people who share some of the diagnostic markers, but will not develop the disease, and those who are true positives for T1D. An important early milestone of the project will be to screen serum samples collected from siblings of T1D index cases who while non-diabetic at enrollment are known to have developed T1D during the course of the longitudinal study. The endpoint of the study is that validated biomarkers will provide sensitive and specific information as to changes in beta cell mass, and/or presence of beta cell autoimmunity, that can be used to inform clinical decisions.

The overall goal of the project is validation of a set of diagnostic proteins to enable monitoring of changes in disease risk. As such, the threshold for success is to establish sensitivity and specificity of the biomarkers. The second step is to demonstrate utility of the agent in the clinic. Clinical utility will require acceptable development of laboratory methods for detecting the selected biomarkers (Weissfeld et al., 1994). These latter studies will be used to show that the protocol can positively impact treatment by identifying at risk individuals and will be the focus of future proposals.

Goal 2. OBJECTIVE AND SPECIFIC AIMS

We propose to work in collaboration with Somalogic, Inc. (a proteomics company based in Boulder, Colorado) to use aptamer array technology to screen samples, serum as well as plasma, collected from the longitudinal cohort of T1D cases and low-risk (non-T1D) cohort in order to identify a set of circulating proteins that can be used to predict disease risk. Using samples available at Children's Hospital of Pittsburgh (CHP), along with data previously generated on the presence of autoantibodies as well as HLA and non-HLA genotypes

associated with T1D susceptibility **we will test the hypothesis that: Circulating Biomarkers can Predict T1D Risk with High Sensitivity and Specificity.**

Goal 2. SPECIFIC AIMS

Aim 1. Using serum collected from converters enrolled prior to onset identify a set of diagnostic biomarkers.

Aim 2. Validate sensitivity of the test to identify converters enrolled in the CHP longitudinal cohort.

Aim 3. Determine specificity of the test to distinguish converters and low-risk individuals enrolled in the CHP longitudinal cohort.

Goal 2. RESEARCH DESIGN AND METHODS:

Aptamer Technology: Aptamers are single-stranded nucleic acids that can be generated to bind tightly to any target molecule (Ellington and Szostak, 1990; Tuerk and Gold, 1990). Development of aptamer technology has achieved antibody like binding properties exemplified by binding kinetics associated with avid binding to cell surface proteins (Ringquist and Parma, 1998), are capable of covalent binding (Jensen et al., 1995), have led to the recognition of naturally occurring riboswitch elements that control gene expression (Nudler and Mironov, 2004), the generation of nucleic acid based drugs for treating ocular disease (Nimjee et al., 2005; Ng et al., 2006), and have been developed as diagnostic tools for quantifying protein abundance in serum as well as plasma (Bock et al., 2004). Aptamer-based multiplexed protein arrays are available from Somalogic and have been validated for the detection of greater than 800 human proteins, including hundreds of cytokines, growth factors, hormones, and cell surface receptors present in circulation. The methodology exhibits picogram per ml sensitivity with 95% coefficients of variation (CV) less than 7% and is capable of precisely quantifying protein abundance over a 5 to 6 log concentration range with sensitivity approaching $1 \times 10^{-15} \text{M}$. The process requires roughly 10 ul plasma or serum for each multiplexed aptamer array, a quantity of material far below what would be required using competing technologies (e.g., Luminex style bead arrays or the SearchLight ELISA platform).

The CHP Longitudinal Cohort: Serum samples are already available from the CHP repository of T1D index cases and their family members. Serum samples have been collected as part of a longitudinal cohort study that includes parents and siblings of T1D index cases (i.e., first degree relatives). The CHP repository dates to 1965 (Libman et al., 1998; Pietropaolo et al., 2005). Approximately 5,000 samples have been collected annually and CHP currently maintains 15 freezers of material and includes 120 to 130 siblings of T1D index cases who entered the study prior to T1D onset. The unique research opportunity that these samples represent is to enable identification of changes in abundance of circulating proteins that can be tested within a highly powered multistage study for the ability to predict disease risk.

Table 1. Selection Criteria.

Aims 1 and 2. Criteria for T1D converters:

- 1) Diagnosed with T1D at 12 years of age or younger.
- 2) Provided multiple serum samples at 12 years of age or younger
- 3) Males and females are equally represented.
- 4) Homozygous for HLA high-risk alleles followed by samples heterozygous for one high-risk allele but do not include any protective alleles.

Aim 3. Criteria for low-risk participants:

- 1) Non-T1D at the endpoint of the longitudinal study.
- 2) Provided multiple serum samples at 12 years of age or younger
- 3) Males and females are equally represented.
- 4) Homozygous for HLA high-risk alleles followed by samples heterozygous for one high-risk allele but do not include any protective alleles.

genotypes. There are anticipated to be a hundred participants who meet the study criteria for T1D conversion and a thousand participants meeting the selection criteria for low-risk (Table 1). The overall selection criteria are expected to reduce confounding due to gender, hormonal changes occurring during adolescence, and genetic risk factors. Selection of a serum sample is contraindicated when CHP records have shown non-T1D

The CHP longitudinal cohort will provide the source of serum samples used during the project. During Aims 1 and 2 samples will be chosen from among individuals recruited prior to T1D onset and will focus on those who developed T1D at age 12 and younger (i.e., prepubescent). Participants chosen for Aim 3 will be selected from among those participants who remained non-T1D at age 18, the study endpoint (Table 1). In both groups of participants (i.e., the T1D converting cohort studied in Aims 1 and 2 as well as the low-risk cohort used during Aim 3) participants will be chosen from those who have provided multiple serum samples. The cohorts will be balanced for males and females. Previously determined genotyping of HLA class II alleles will be used to enable the selection of participants homozygous for HLA susceptibility alleles followed by participants heterozygous for one susceptibility allele but will not include any participants heterozygous for HLA protective

related illness occurring 2 weeks prior to the blood draw. Participation of an individual is contraindicated when CHP records show frequent, unexplained non-T1D related illness that exceed the 95th percentile of all participants.

Study Design and Data Analysis: The research plan is to use the CHP repository of serum samples to provide materials that can be evaluated for the abundance of >800 different protein species. During Aim 1 of the project researchers at CHP will work in collaboration with Somalogic to choose samples that meet the selection criteria (Table 1) from a sufficient number of participants to establish temporal trends as well as age dependent reference intervals for serum proteins prior to T1D onset. A milestone of the study is to obtain measurements 4 times (but no less than 3 times) on each serum sample. It is anticipated that as many as 3 serum samples will be analyzed for each participant. A single sample will correspond as closely as possible to a time point immediately after T1D onset while the remaining samples will correspond to times preceding T1D diagnosis by 1 and 5 years. The resulting dataset will enable a within person study that will be analyzed to identify those proteins that deviate from the normal range of abundance as the disease event approaches.

During Aim 2 we will use the results garnered during the preceding Aim to validate the sensitivity of selected proteins to forecast T1D onset. This will be performed using same cohort of T1D converter participants chosen from the CHP longitudinal cohort (Table 1). The study will employ additional serum samples collected at 2 additional time points (e.g., preceding onset by 3 and 7 years) from the same participants examined during Aim 1. The goal of Aim 2 of the study is to extend the time series in order to determine the greatest time prior to disease event in which to validate sensitivity of serum markers for association with T1D.

Aim 3 of the study can be performed simultaneous with Aim 2 and will test the specificity of the biomarkers to forecast T1D onset. Serum samples will be chosen as summarized in Table 1 using material collected from siblings of T1D cases that remained non-T1D at the study endpoint (i.e., low-risk subjects under 18 years of age). Analysis of the abundance of select serum proteins will be used to establish age dependent reference ranges for the low-risk cohort. Comparison of the mean abundance and variance associated with individual protein species measured using the T1D converting cohort with data generated from the low-risk cohort will be used to estimate the overall sensitivity of the biomarkers for forecasting disease.

Table 2. Power Analysis.

<u>Study Aim</u>	<u>Cohort</u>	<u>Size</u>
1 and 2	Converters	73
3	Low-Risk	657

Power Analysis: A preliminary power analysis was performed in order to determine the sample size in the overall diagnostic study. The approach employed the calculations provided by Jones et al. (2003) and estimated the number of participants necessary to achieve 0.95 sensitivity and 0.80 specificity for p-values of 0.05 and 80% power. The calculation assumes the

prevalence of T1D in the selected population will be 0.10 (i.e., using numbers of T1D converters and low-risk subjects to approximate the prevalence of T1D among siblings of index cases) and that the confidence interval will be equivalent to 0.05. In this manner the estimated total number of study participants required for determining whether selected biomarkers are sensitive and specific is 730. The experimental plan for incorporating the participants into the proposed study is summarized in Table 2.

The nomogram provided by Altman (1991) was used in a second power analysis in order to estimate the sample size needed to achieve 90% power and p-value of 0.01 when samples collected within and between persons are analyzed. The calculation assumes that the standard deviation in the data is roughly 2-fold within subjects (Anderson and Anderson, 2003) while the deviation observed between subjects may be as large as 4-fold. During analysis of data obtained from within person studies (N=73) it is estimated that the approach will exhibit 90% power to detect a 0.8-fold change in protein abundance at a p-value of 0.01. During comparison between subjects the study will achieve a significance of 0.01 and 90% power to detect a 1.9-fold change in protein abundance when 73 converters and 657 low risk subjects are compared. The approach does not account for confounding due to multiple testing which can dramatically increase the required number of subjects. For example, the research will test >800 protein species. Therefore, the number of independent tests will influence the threshold for significance and is estimated to be 1×10^{-5} (i.e., 0.01/800). It is possible that the results garnered during Aim 1 may fail to achieve the corrected threshold for significance. If this is the case then proteins will be sorted based upon the observed confidence intervals and uncorrected p-values. Proteins exhibiting the smallest confidence intervals and p-values will be chosen for study during Aims 2 and 3.

Publications and Grant Applications: For researchers at CHP the primary outcome of the study will be to use the data in peer-reviewed publications and in presentations at scientific meetings. The CHP group will also use the data in support of grant applications. We will work together with Somalogic to achieve these goals, however, the first and corresponding authors on publications resulting from the collaboration will be affiliated with the Pittsburgh team.

Goal 2. TIMELINE: but timeless for now

Aim 1: CHP researchers in consultation with Somalogic will choose a set of serum samples collected at multiple time points from T1D cases. Blinded samples will be shipped to Somalogic for testing on aptamer arrays and quality control analysis of the resulting data. Following quality control analysis the samples will be unblinded. CHP and Somalogic will share raw as well as analyzed data with the goal of identifying the optimal set of serum biomarkers exhibiting sensitivity to T1D risk.

Aim 2: CHP researchers and Somalogic will work together to select a second set of serum samples from T1D cases that can be used to validate markers discovered during Aim 1 of the project. Blinded samples will be shipped to Somalogic for testing and will be unblinded following quality control analysis. In this example only the biomarkers identified during Aim 1 as being sensitive for T1D risk need to be analyzed. CHP and Somalogic will share in the data analysis steps.

Aim 3: Samples collected from low-risk participants can be analyzed simultaneously with the experiments described in Aim 2 of the project. Scientists at CHP and Somalogic will work together to select a set of serum samples from low-risk individuals recruited in the CHP longitudinal cohort. These samples have been collected at multiple time points from participants who have remained non-T1D at the study's endpoint. Blinded samples will be used and following quality control steps will be unblinded. Data analysis steps will be shared between CHP and Somalogic.

Goal 2. LITERATURE CITED

Altman DG. (1991) In: *Practical Statistics for Medical Research*. Chapman & Hall/CRC, Boca Raton, FL.

Anderson NL, Anderson NG. (2003) The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics* 1:845-867.

Barker JM, Eisenbarth GS. (2004) The natural history of autoimmunity in type 1a diabetes mellitus. In: *Diabetes Mellitus: A Fundamental and Clinical Text 3rd Edition*. LeRoith D, Taylor SI, Olefsky JM eds. Lippincott, Williams, and Wilkins, Philadelphia, PA, pp. 471-482.

Bock C, Coleman M, Collins B, Davis J, Foulds G, Gold L, Greef C, Heil J, Heilig JS, Hicke B, Hurst MN, Husar GM, Miller D, Ostroff R, Petach H, Schneider D, Vant-Hull B, Waugh S, Weiss A, Wilcox SK, Zichi D. (2004) Photoaptamer arrays applied to multiplexed proteomic analysis. *Proteomics* 4:609-618.

Butler PC, Meier JJ, Butler AE, Bhushan A. (2007) The replication of beta cells in normal physiology, in disease and for therapy. *Nat Clin Pract Endocrinol Metab* 3:758-768.

Ellington AD, Szostak JW. (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature* 346:818-822.

Filippi CM, von Herrath MG. (2008) Viral trigger for type 1 diabetes: pros and cons. *Diabetes* 57:2863-2871.

Jensen KB, Atkinson BL, Willis MC, Koch TH, Gold L. (1995) Using in vitro selection to direct the covalent attachment of human immunodeficiency virus type 1 Rev protein to high-affinity RNA ligands. *Proc Natl Acad Sci USA* 92:12220-12224.

Jones SR, Carley S, Harrison M. (2003) An introduction to power and sample size estimation. *Emerg Med J*. 20:453-458.

- Karvonen M, Viik-Kajander M, Moltchanova E, Libman I, LaPorte R, Tuomilehto J. (2000) Incidence of childhood type 1 diabetes worldwide. Diabetes Mondiale (DiaMond) Project Group. Diabetes Care 23:1516-1526.
- Klitz W, Maiers M, Spellman S, Baxter-Lowe LA, Schneckpeper B, Williams TM, Fernandez-Viña M. (2003) New HLA haplotype frequency reference standards: high-resolution and large sample typing of HLA DR-DQ haplotypes in a sample of European Americans. Tissue Antigens 62:296-307.
- Libman IM, LaPorte RE, Becker D, Dorman JS, Drash AL, Kuller L. (1998) Was there an epidemic of diabetes in nonwhite adolescents in Allegheny County, Pennsylvania? Diabetes Care 21:1278-1281.
- Newacheck PW, Taylor WR. (1992) Childhood chronic illness: prevalence, severity, and impact. Am J Public Health 82:364-371.
- Ng EW, Shima DT, Calias P, Cunningham ET Jr, Guyer DR, Adamis AP. (2006) Pegaptanib, a targeted anti-VEGF aptamer for ocular vascular disease. Nat Rev Drug Discov 5:123-132.
- Nimjee SM, Rusconi CP, Sullenger BA. (2005) Aptamers: an emerging class of therapeutics. Annu Rev Med 56:555-583.
- Nudler E, Mironov AS. (2004) The riboswitch control of bacterial metabolism. Trends Biochem Sci 29:11-17.
- Palmer JP, Fleming GA, Greenbaum CJ, Herold KC, Jansa LD, Kolb H, Lachin JM, Polonsky KS, Pozzilli P, Skyler JS, Steffes MW. (2004) C-peptide is the appropriate outcome measure for type 1 diabetes clinical trials to preserve beta-cell function: report of an ADA workshop, 21-22 October 2001. Diabetes 53:250-264.
- Pasquali L, Fan Y, Trucco M, Ringquist S. (2006) Rehabilitation of adaptive immunity and regeneration of beta cells. Trends Biotechnol 24:516-522.
- Pietropaolo M, Yu S, Libman IM, Pietropaolo SL, Riley K, LaPorte RE, Drash AL, Mazumdar S, Trucco M, Becker DJ. (2005) Cytoplasmic islet cell antibodies remain valuable in defining risk of progression to type 1 diabetes in subjects with other islet autoantibodies. Pediatr Diabetes 6:184-192.
- Pihoker C, Gilliam LK, Hampe CS, Lernmark A. (2005) Autoantibodies in diabetes. Diabetes 54(Suppl 2):S52-61.
- Redondo MJ, Jeffrey J, Fain PR, Eisenbarth GS, Orban T. (2008) Concordance for islet autoimmunity among monozygotic twins. N Engl J Med 359:2849-2850.
- Rewers A, Chase HP, Mackenzie T, Walravens P, Roback M, Rewers M, Hamman RF, Klingensmith G. (2002) Predictors of acute complications in children with type 1 diabetes. JAMA 287:2511-2518.
- Riley WJ, Maclaren NK, Krischer J, Spillar RP, Silverstein JH, Schatz DA, Schwartz S, Malone J, Shah S, Vadheim C, et al. (1990) A prospective study of the development of diabetes in relatives of patients with insulin-dependent diabetes. N Engl J Med 323:1167-1172.
- Ringquist S, Parma D. (1998) Anti-L-selectin oligonucleotide ligands recognize CD62L-positive leukocytes: binding affinity and specificity of univalent and bivalent ligands. Cytometry 33:394-405.
- Steele C, Hagopian WA, Gitelman S, Masharani U, Cavaghan M, Rother KI, Donaldson D, Harlan DM, Bluestone J, Herold KC. (2004) Insulin secretion in type 1 diabetes. Diabetes 53:426-433.
- Trucco M. (2005) Regeneration of the pancreatic beta cell. J Clin Invest 115:5-12.

Tuerk C, Gold L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* 249:505-510.

Weissfeld JL, Larsen RD, Niman HL, Kuller LH. (1994) Evaluation of oncogene-related proteins in serum. *Cancer Epidemiol Biomarkers Prev* 3:57-62.

Verge CF, Gianani R, Kawasaki E, Yu L, Pietropaolo M, Jackson RA, Chase HP, Eisenbarth GS. (1996) Prediction of type I diabetes in first-degree relatives using a combination of insulin, GAD, and ICA512bdc/IA-2 autoantibodies. *Diabetes* 45:926-933.

Statement of Plans for the Upcoming Research Period

Goal 1. Extend genetic evaluation of T1DN by incorporating the results from GWAS published in the scientific literature. **Milestone 1A.** Incorporate GWAS data gathered by GAIN. **Milestone 1B.** Incorporate GWAS of data garnered by DCCT/EDIC.

Goal 2. Using the results from the combined analysis of Children's Hospital of Pittsburgh and publicly available data choose genomic regions for fine mapping. **Milestone 2A.** Select regions for fine mapping. **Milestone 2B.** Select tag-SNPs that will be genotyped. **Milestone 2C.** Design experimental assays for laboratory analysis.

Goal 3. Initiate fine mapping of T1DN susceptibility regions. **Milestone 3A.** Genotype selected SNPs in T1DN cases and T1D controls. **Milestone 3B.** Analyze results to determine association between variant(s) and T1DN phenotype.

In the fourth and final quarterly scientific progress report (06/01/09 - 08/26/09) of year 02, we now report on our cumulative results.

The goals of the recently concluded research quarter have been to: 1) extend the genetic evaluation of T1DN by incorporating results from published GWAS data; 2) use the results from the combined analysis of Children's Hospital of Pittsburgh and publicly available data to choose genomic regions for fine mapping; and 3) initiate fine mapping of T1DN susceptibility regions. The quarterly goals 1 and 2 have been completed while the research to complete goal 3 is ongoing.

Goal 1. *Extend genetic evaluation of T1DN by incorporating the results from GWAS published in the scientific literature.* **Milestone 1A.** *Incorporate GWAS data gathered by GAIN.* **Milestone 1B.** *Incorporate GWAS of data garnered by DCCT/EDIC.*

Completion of Milestones 1A and 1B: Incorporate GWAS data gathered by GAIN and incorporate GWAS of data garnered by DCCT/EDIC. Data from the GAIN study of T1DN in the GoKinD and DCCT cohorts have been published (Pezzolesi et al., 2009). The combined analysis presented results similar to our own study of T1DN in the GoKinD cohort (data presented in our previous Quarterly Reports). For example, the strongest 11 SNPs identified 4 genomic regions as being of potential association with the phenotype (Table 1). The region on Chromosome 13q was identified in our work as well as in the results published by the GAIN consortium. Additional genetic markers were also reported by the joint GoKinD and DCCT analysis. These included Chromosome 7p (*CPVL*, *CHN2*), Chromosome 9q (*FRMD3*), and 11p (*CARS*).

Table 1. Comparison of data obtained from the CHP and GAIN studies on the genetics of T1DN.

SNP	Chr	Position	Nearest Gene(s)	Values Reported GAIN		Values Reported CHP	
				p-value	OR(95% CI)	p-value	OR(95% CI)
rs39059	7p	29.2	<i>CPVL/CHN2</i>	5.0E-06	1.4(1.2-1.6)	----	----
rs39075	7p	29.2	<i>CPVL/CHN2</i>	6.5E-07	1.4(1.2-1.6)	----	----
rs1888747	9q	85.3	<i>FRMD3</i>	6.3E-07	1.5(1.3-1.7)	----	----
rs10868025	9q	85.4	<i>FRMD3</i>	5.0E-07	1.5(1.3-1.7)	----	----

rs451041	11p	3.0	CARS	3.1E-06	1.4(1.2-1.6)	----	----
rs739401	11p	3.0	CARS	6.4E-06	1.4(1.2-1.6)	----	----
rs1041466	13q	109.0	IRS2	3.2E-06	1.4(1.2-1.6)	2.9E-06	1.4(1.2-1.6)
rs2150479	13q	109.0	IRS2	----	----	2.6E-06	1.5(1.3-1.7)
rs1411766 & rs17412858	13q	109.1	IRS2	1.8E-06	1.4(1.2-1.6)	1.8E-05	1.4(1.2-1.6)
rs6492208 & rs2391777	13q	109.1	IRS2	6.1E-06	1.4(1.2-1.6)	1.9E-05	1.4(1.2-1.6)
rs7989848 & rs7989418	13q	109.1	IRS2	7.0E-06	1.4(1.2-1.6)	1.4E-05	1.4(1.2-1.6)

The p-values for association of these regions with T1DN phenotype ranged between 7×10^{-6} and 5×10^{-7} . Well above the threshold for genome-wide significance assigned at 1.4×10^{-7} for this study. A likely interpretation of the data is that while family inheritance of T1DN has been estimated at 2 to 3-fold the markers assayed by the GWAS represented alleles that were in linkage disequilibrium with causal genes but were not themselves causal. Alternatively, shared environmental influences confounded estimation of the genetic risk. In other words, it is the combined influence of shared genes and gene-environment interaction that results in 2 to 3-fold risk among siblings. Therefore, the signals identified (and reported previously in our quarterly reports) may in fact be consistent with the anticipated signals.

What does this mean for future work in this field? The balance between strength of the genetic effect, minor allele frequency, determine the sample size required to achieve a p-value that is significant in the face of the multiple testing error. Estimations of sample size are summarized in Table 2. For the odds ratio of 1.4 that was observed by GAIN as well as by our own analysis the expected number of required participants varied between 2,000 and 4,800 for minor allele frequencies greater ranging between 0.5 and 0.1. This is more than the roughly 1,700 subjects have been studied so far. At this point, a way forward is to identify additional subjects for genetic analysis.

Table 2. Number of subjects to achieve genome-wide significance.

MAF	Odds Ratio	Number of Subjects to Genotype
0.10	1.4	4,800
0.20	1.4	2,810
0.30	1.4	2,220
0.40	1.4	2,010
0.50	1.4	2,000

Goal 2. Using the results from the combined analysis of Children's Hospital of Pittsburgh and publicly available data choose genomic regions for fine mapping. **Milestone 2A.** Select regions for fine mapping. **Milestone 2B.** Select tag-SNPs that will be genotyped. **Milestone 2C.** Design experimental assays for laboratory analysis.

Completion of Milestones 2A, 2B, and 2C: Select regions for fine mapping, select tag-SNPs that will be genotyped, and design experimental assays for laboratory analysis. In order to accomplish this goal we have evaluated the combined data obtained from CHP and GAIN. The results have been described in the section summarizing the results of Goal 1 and indicated that no SNP evaluated by either group achieved genome-wide significance. As a result, we have modified the Goals 2 and 3 of the recently completed research quarter. The new Goal has been to quickly identify a new cohort of T1DN subjects. This is being accomplished by discussions that are ongoing between our group and the group lead by our collaborator Bernhard Boehm. Together we have identified potentially suitable cohorts that have been recruited for T1D and T1DN studies. The cohort has been recruited from the DPV-registry of German T1D patients and consists of greater than 14,000 subjects. There are plasma, DNA, and relevant medical records available. We estimate that 10-30% of the T1D subjects will develop T1DN. The experimental plan that is being discussed is to measure serum Cystatin C levels to estimate each subjects T1DN status and to use T1DN positive patients to supplement our collection of samples obtained from GoKinD and the CHP recruitment.

At present our estimated cohorts for the T1DN study consist of N=835 of cases and N=908 of controls (Table 3). Use of the DPV cohort is anticipated to increase the case cohort to as many as N=5,105 cases and N=10,708 controls. This will correspond to a number of cases that are estimated to be sufficient to evaluate T1DN risk alleles identified by data reported from our previously reported research efforts.

Table 3. T1DN cohort.

<u>Cohort</u>	<u>T1DN Cases</u>	<u>T1D Controls</u>
Genetics of Kinneys in Diabetes	835	908
Children's Hospital of Pittsburgh	70	----
DVV-Registry (Estimated)	4,200	9,800
Total Number of Subjects	5,105	10,708

Goal 3. *Initiate fine mapping of T1DN susceptibility regions.* **Milestone 3A.** *Genotype selected SNPs in T1DN cases and T1D controls.* **Milestone 3B.** *Analyze results to determine association between variant(s) and T1DN phenotype.*

Progress towards the completion of Milestones 3A and 3B: Genotype selected SNPs in T1DN cases and T1D controls as well as analyze results to determine association between variants(s) and T1DN phenotype. The Goal 3 will be pursued along with our efforts to obtain additional T1DN cases and controls. Collection of additional samples to provide the cohort sizes needed to achieve genome-wide significance will precede genetic analysis. We anticipated that this work should account for a significant amount of effort during the no-cost extension that has been requested for the project.

Literature Cited:

Pezzolesi MG, Poznik GD, Mychaleckyj JC, Paterson AD, Barati MT, Klein JB, Ng DP, Placha G, Canani LH, Bochenski J, Waggott D, Merchant ML, Krolewski B, Mirea L, Wanik K, Katavetin P, Kure M, Wolkow P, Dunn JS, Smiles A, Walker WH, Boright AP, Bull SB; DCCT/EDIC Research Group, Doria A, Rogus JJ, Rich SS, Warram JH, Krolewski AS. (2009) Genome-wide association scan for diabetic nephropathy susceptibility genes in type 1 diabetes. *Diabetes* 58:1403-1410.

KEY RESEARCH ACCOMPLISHMENTS:

1. Creation of a sample repository containing DNA from 905 T1DN and 908 non-T1D participants.
2. Establishment of laboratory methods for multiplex genotyping of 30 or more SNPs simultaneously from as many as 5,000 samples.
3. Initiation of genetic studies to evaluate candidate genetic markers for association with diabetes complications.
4. Analysis of genetic data to evaluate candidate genes for association with diabetes phenotypes.
5. Identification of additional T1DN cohorts available for genetic analysis of candidate risk alleles.
6. Publication of 10 manuscripts.

REPORTABLE OUTCOMES:

Manuscripts (2 publications)

1. Lu, L., Boehm, J., Nichol, L., Trucco, M., and Ringquist, S. Multiplex HLA typing by pyrosequencing. In *Methods in Molecular Biology*, vol 496: DNA and RNA Profiling in Human Blood. ed. P. Bugert. Humana Press Inc., Totowa, New Jersey (2009).

2. Wu J, Devlin B, Ringquist S, Trucco M, and Roeder K. Screen and clean: a tool for identifying interactions in genome-wide association studies. Genetic Epidemiology, Submitted, 2009.

Abstracts

None

Presentations

None

Patents and Licenses Applied for and/or Issued

None

Degrees Obtained that are Supported by this Award

None

Development of Cell Lines, Tissue or Serum Repositories

1. Repository of DNA samples collected from T1D and T1DN patients exceeding 1,800 subjects.

Informatics such as Databases and Animal Models, etc

None

Funding Applied for Based on Work Supported by this Award

1. An application for funding has been submitted to the National Institutes of Health in response to RFA-DK-08-006 Fine Mapping and Function of Genes for Type 1 Diabetes. The project is entitled "Refining the Genetic and Functional Architecture of Type 1 Diabetes" and the goal is to elucidate molecular networks affecting T1D susceptibility that are directly influenced by stably inherited genetic variants. Increased understanding of the gene-networks underlying disease risk will aid in the development of accurate screening tools as well as in creation of new therapeutic treatments.

Employment or Research Opportunities Applied for and/or Received Based on Experience/Training Supported by this Award

None

CONCLUSION:

The conclusions for the first and second year of funding are that statistically significant genetic markers associated with T1DN have been identified on Chromosomes 7p, 9q, 11p, and 13q. Comparison of the p-values for these markers indicate that additional subjects with T1DN (cases) and T1D (controls) need to be recruited in order to replicate the findings and to establish statistical significance exceeding the threshold of genome-wide significance. A second cohort of N=14,000 T1D subjects has been identified. Work that is planned for the upcoming year (covered by the requested no-cost extension) will seek to evaluate the new cohort of T1D subjects for evidence of T1DN as well as genotyping of select T1DN risk loci.

The research project generated 10 publications. These are listed under the section entitled "REPORTABLE OUTCOMES".

The So What Section. What are the implications of this research? Diabetes affects 16 million Americans and 800,000 new cases annually. African, Hispanic, Native and Asian Americans are particularly susceptible to its most severe complications. Costs associated with diabetes may be as high as \$132 billion. Diabetes accounts for 42% of new cases of end-stage renal disease with over new 100,000 cases per year at an average cost of \$55,000 per patient annually.

What are the military significance and public purpose of this research? As the military is a reflection of the U.S. population improved prediction of risk for developing diabetes and diabetic complications among active duty

members of the military, their families, and retired military personnel will potentially allow focused preventative treatment of at risk individuals, providing significant healthcare savings and improved patient well being.

REFERENCES:

1. Ringquist, S, Styche, A., Rudert, W.A., and Trucco, M. Pyrosequence-based strategies for improved allele typing of HLA loci. In: *Methods in Molecular Biology*, vol 373: *Pyrosequencing Protocols*. ed. S. March. Humana Press Inc., Totowa, New Jersey (2007).
2. Ringquist, S., Pecoraro, C., Lu, Y., Styche, A., Rudert, W.A., Benos, P.V., and Trucco, M. Web-based primer design software for genome scale SNP mapping by pyrosequencing. In: *Methods in Molecular Biology*, vol 373: *Pyrosequencing Protocols*. ed. S. March. Humana Press Inc., Totowa, New Jersey (2007).
3. Ringquist, S., Nichol, L., and Trucco, M. Transplantation Genetics. In *Emery and Rimoin's Principles and Practice of Medical Genetics*. eds. D.L. Rimoin, M. Conner, R.E. Pyeritz, B.R. Korf, and A.E. Emery 5th Edition, Elsevier Books, Oxford (2007).
4. Pasquali, L., Bhargava, R., Ringquist, S., Styche, A., Bedeir, A., and Trucco, G. Quantitative methylation of CpG islands in progressive breast neoplastic lesions from normal to invasive carcinoma. *Cancer Letters* 257, 136-144 (2007).
5. Ringquist, S., Pecoraro, C., and Trucco, M. Web-based program for pyrosequencing primer design. *ASHI Quarterly* 31, 50-52 (2007).
6. Pasquali, L., Trucco, M., and Ringquist, S. Navigating pathways affecting type 1 diabetic kidney disease. *Pediatric Diabetes* 8, 307-322 (2007).
7. Luca, D., Ringquist, S., Klei, L., Lee, A.B., Gieger, C., Wichmann, H.-E., Schreiber, S., Krawczak, M., Lu, Y., Styche, A., Devlin, B., Roeder, K., and Trucco, M. On the use of general control samples for genome-wide association scans: genetic matching highlights causal variants. *American Journal of Human Genetics* 82, 452-463 (2008).
8. Zhang, L., Ringquist, S., Perdomo, G., Qu, S., Trucco, M., and Dong, H.H. Proteomic analysis of fructose-induced fatty liver in hamsters. *Metabolism* 57, 1115-1124 (2008).
9. Lu, L., Boehm, J., Nichol, L., Trucco, M., and Ringquist, S. Multiplex HLA typing by pyrosequencing. In *Methods in Molecular Biology*, vol 496: *DNA and RNA Profiling in Human Blood*. ed. P. Bugert. Humana Press Inc., Totowa, New Jersey (2009).
10. Wu J, Devlin B, Ringquist S, Trucco M, and Roeder K. Screen and clean: a tool for identifying interactions in genome-wide association studies. *Genetic Epidemiology*, Submitted, 2009.

APPENDICES: None

SUPPORTING DATA: None